

# Genetic Evidence of Early Human Migrations in the Indian Ocean Region Disproves Aryan Migration/ Invasion Theories

An Examination of Small-statured Human Groups of the Indian Ocean Region

*Lavanya Vemsani*

THE theory of Indo-Aryans (although no race or tribe exists by that name), and their migration/invasion into India is based on scant linguistic and anthropological evidence. Historically, physical attributes or cultural attributes such as language are used to differentiate humanity into various ethnic human groups. Initially, differences in physical characteristics and attributes such as height, weight, shape of face, eyes and nose were seen as notable characteristics in determining one's ethnicity in the eighteenth and early nineteenth centuries ce. However, this quickly gave way to more sophisticated theories based on linguistics, as a tool to understand population migrations based on historical analysis of languages spoken by different groups across large areas although they might not exactly share same physical characteristics. Research based on these two tools have not solved the issue of Indo-Aryans, but led to lengthy debates among the scholars proposing opposing viewpoints, thereby confusing whole scenario of the origin and spread of Indo-Aryan languages sometimes linked to population migration and sometimes not linked to such migrations.<sup>1</sup> Although some middle ground is established lately between these two camps (for and against Indo-Aryan invasions/migration theories), linguistics and anthropological or archaeological research needs additional resources to understand or solve this issue completely. The only way out of this gridlock is to use a new parameter that provides undisputable scientific evidence, such as the genetic research.<sup>2</sup>

<sup>1</sup> Bryant, Edwin and Laurie L. Patton, 2005; Bryant, E. 2001.

<sup>2</sup> Lee et al. 2008.

The human genome mapping is carried on in a big way throughout the world with a number of resource and data networks created for the purpose of disseminating knowledge (<https://genographic.nationalgeographic.com/genographic/index.html>; [www.isogg.org](http://www.isogg.org); [http://www.ornl.gov/sci/techresources/Human\\_Genome/home.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml)). This project provided evidence to locate the origin of various human groups, and their spread across the world, through a long span of time.<sup>3</sup> This genetic map not only establishes that no new genetically diverse groups entered India after the first settlements from 80 to 50 kya (thousand years ago), but that the descendants of this first genetic group, which populated the subcontinent also spread to other parts of the world. These landmark findings not only propose new data on human origins, but also supplement the findings of physical anthropology, and add new dimensions to understand population migrations. It is only pertinent that historians take advantage of this new data and apply it to some of the controversies of history.

---

→ Behar et al. 2012. The article provides a reconstructed sequence of the largest available resource of mtDNA with a view to have a substantial practical and educational influence on the scientific and public perception of human evolution by clarifying the core principles of common ancestry for extant descendants of the world.

Endicott et al. 2007. This article provides a summary of Pleistocene settlements of early humans in India.

Underhill et al. 2000, "Y-chromosome Sequence Variation and the History of Human Populations", *Nature Genetics*, 26.3: 358-61.

Genetic evidence of biota is also under scrutiny to understand the dispersals of animal and plant species across the continents through marine dispersal. The following article discusses the possibility of India as a central place of dispersal for a number of animal and plant species. See A. Datta-Roy and Praveen Karanth, 2009, "The Out of India Hypothesis: What Do Molecules Suggest?", *Journal of Biosciences*, 34(5): 687-97.

<sup>3</sup> The human genome holds genetic information encoded in the coiled DNA of the chromosomes, which is an immense document written in four letters — G, C, A, T (guanine, cytosine, adenine and thymine). Both X and Y-chromosomes are sex-specific chromosomes, and haploid (that is have only one combination of X or Y, with 23 chromosomes), while all the other cells have 46 chromosomes. The human Y-chromosome consists of a Non-Recombining Region (NRY), making up 95 per cent of its length, flanked by autosomal regions, which is passed intact without change in the male-line except for mutations. While mtDNA is a female-specific passed through the female line from mother to daughter without change except for mutations, therefore permits tracking female heritage. Y is a male-specific chromosome passed from father to son. Binary polymorphisms of the non-recombining region of the human Y-chromosome preserves the paternal genetic legacy of humans that has persisted up to the present. A glossary of terms related to the genome can be found at: <https://genographic.nationalgeographic.com/genographic/lan/en/glossary.html>

In this paper I use the available genetic research data to analyse the genetic history of India with a view to understand the historic peopling of India. This paper is divided into two parts, in the first part, I will examine the genetic history of India from female genetic heritage, and propose that the original settlers not only continued to propagate in India but spread out on to Australia.<sup>4</sup> In the second part I examine the male genetic heritage in India, and how it is uniquely shared with certain groups of populations known as relict populations (variously termed as Negroid, Mongoloid or Australoid groups), and thereby provide evidence of the spread of early humans out of India. This study therefore demonstrates that the female genetic lines spread from India to Australia and then to other parts of the world, while male genetic lines spread uniquely from India to South-East Asia, and then on to Europe and America.<sup>5</sup>

#### **Short-statured Human Groups of Indian Ocean Region: Pioneers and Founders of Non-African Lineages Across the World**

The various human groups of the world evolved from a common small hominid type that originated in Africa, spread around the Indian Ocean rim, where pockets of peoples of small stature still continue to live. Although the similarities in the physical features of these isolated groups of small-statured people are noted with the common nomenclature given to them (Negroid, Australoid, Veddoid, etc.) in the Indian Ocean region, it was not considered until recently that they may all have been derived from a common ancestor. However, recent mitochondrial and Y-chromosome analysis of these small-statured, isolated populations in the Indian Ocean region in fact established that these isolated groups are the relict population left from the late Pleistocene (130-10 KYA) and represent the migration of anatomically modern humans (*Homo sapiens sapiens*) from Africa, which reached Australia by Holocene (50-10 KYA) and then elsewhere in the world. In this great Pleistocene migration India is at the crossroads, and acted as a premier corridor of humans migrations towards East and West.<sup>6</sup>

<sup>4</sup> Mountain et al. 1995, "Demographic History of India and mtDNA-sequence Diversity", *Am. J. Hum. Genet.*, 56.4: 979-92. This article is one of the earliest attempts to reconstruct the population genetics of India and proposes an indigenous origin of Indian human groups. Further research also showed indigenous development and positive selection amidst notable diversity in the genetic history of India (Metspalu et al. 2011).

<sup>5</sup> Kivisild et al. 1999b.

Douglas Wallace and Antonio Torroni, 2009, "American Indian Prehistory as Written in the Mitochondrial DNA: A Review", *Human Biology*, 81(5/6) (October): 509-21.

Karafet et al. 1999; Torroni et al. 2001.

<sup>6</sup> Endicott et al. 2007.

The Indian Ocean region consists of East Coast of Africa, Arab Peninsula, South Asia (India, Pakistan, Sri Lanka and Bangladesh), South-East Asia (Indonesia, Thailand, Philippines, Java, Borneo and other islands), Malaysia (Peninsular and Island Malaysia) and Australia (north and western coast). Since east coast of Africa contains Africa-specific DNA lineages and the Arab Peninsula contains very young DNA clades of non-African heritage, I will limit my analysis here to South Asian, South-East Asian regions in this paper. Many of the small-statured groups in India and South-East Asia are considered tribal populations, designated by various names, derogatory or otherwise, and have never been consistently studied or understood precisely. Numerous racial theories have been proposed to classify them together as one race dispersed across the Indian Ocean or as numerous races completely different from each other. Existence of numerous languages and diverse physical features amidst a few similarities compounded the racial theories, resulting in contradictory propositions.<sup>7</sup> Contrary to the early assumptions, the small-statured human groups scattered across the Indian Ocean region do not constitute a single racial or linguistic entity, but an ancestral group, migrated from Africa during the Pleistocene era, and still hold the most ancient DNA of the region in which they are located, and hence can be considered the ancestors of all non-African humanity on the earth, including India. Discussed below are the geographical location, ethnic and DNA profiles of these ancestral populations in the Indian Ocean region.

#### FEMALE GENETIC PROFILES OF THE INDIAN OCEAN REGION

##### *India*

India is the largest country in the Indian Ocean region (seventh largest in the world) with the largest population (over 1 billion) and possesses the greatest number of ethnic and linguistic groups. India measures 3,214 km (1,997 miles) from north to south and 2,993 km (1,860 miles) from east to west with a coastline of 7,516.5 km (4,671 miles).<sup>8</sup> This coastline has been utilized since Pleistocene era for living as well as travel and making contacts with the surrounding regions, as noticed from the coastal migrations brought in to light by the genetic research.<sup>9</sup>

<sup>7</sup> Bulbeck et al. 2006. This paper strongly demonstrates the futility of classifying humanity into races based on differences on measurements of various physical features of the body of humans.

<sup>8</sup> Data based on the Research, Reference and Training Division, Government of India official web portal <http://www.india.gov.in/knowindia/profile.php>[accessed 8 July 2012].

<sup>9</sup> Although there are differing opinions concerning the mode of travel of these early settlers,

India consists of a large number of social groups classed as tribal groups, known as *ādivāsīs* (Sanskrit: first residents) now categorized by the Government of India, with a special constitutional status as the Scheduled Tribes and makes up about 6.5 per cent of India's population ([http://www.india.gov.in/govt/constitutions\\_india.php](http://www.india.gov.in/govt/constitutions_india.php)). These tribal groups are endogamous and are spread across Indian subcontinent and islands, speak various languages and exhibit distinct physical characteristics, although grouped under a single designation, as "tribe", this is the most diverse group of Indian population,<sup>10</sup> the only commonality among them being lifestyle practices, which derive from prehistoric hunting-gathering and pastoralist lifestyles.<sup>11</sup> The major tribal groups of north and north-west India are: Lambadas (Rajasthan), Kolis, Mukris, Yanadi, Erukala (Karnataka). The major tribal groups of north-east India are: Angāmī, Nāgā, Kuki, Lodhā, Khariā, Mundaři; the major tribal groups of south-west India are Panyar, Yanadi; the major tribal groups of south-east India are: Chenchus, Koyās, Koṇḍs, Muṇḍarīs, Aṇḍamānes, Orīge, Jorawa, Shompen and Sentinelese of Andaman Nicobar islands. All of these groups sustain on a hunter-gatherer lifestyle with intermittent swidden (slash and burn) agriculture until recently, when, due to government efforts, they are adopting to somewhat more settled life practices. Tribes of India have attracted major focus in the renewed interest in understanding human origin and spread through genetic evidence and its relation to languages.<sup>12</sup> However, research established that the numerous endogamous caste groups of India (Hindu *jātis*) do not differ greatly with regard to genetics from their tribal neighbours.<sup>13</sup>

#### *Female Genetic Configuration of Indians*

All non-Africans have inherited a subset of L3 ancestral line out of Africa, mtDNA lineages, differentiated into haplogroups M and N. Numerous studies are published on the genetic profiles of Indians including tribal and caste populations, with very old mtDNA (female) dated for M clade, while N and R clades have originated in India itself.<sup>14</sup> Scholars differed initially on the place of origin of M to be located either in

---

→ a consensus recently supports the coastal migration route of the early Pleistocene settlers across the Indian Ocean rim.

Macaulay et al. 2005; David Bulbeck 2007; Metspalu et al. 2011: 744.

<sup>10</sup> Watkins et al. 2005.

<sup>11</sup> Singh 1998.

<sup>12</sup> Chauhan et al. 2011; Chauhan et al. 2008b.

<sup>13</sup> Chauhan et al. 2007.

<sup>14</sup> Palanichamy et al. 2004.

Africa or India, while the most recent studies lean towards placing its origin in India, due to the lack of any ancient and deep-rooted mitochondrial M clades in Africa or West Asia, except for a single, very young M1 clade in Ethiopia, which is considered to be the result of a back migration.<sup>15</sup> While India possesses very ancient and deep-rooting mtDNA for M, N and R, West Asia and Europe only have younger M derivative lineages, and South-East Asia and East Asia have younger N derivative lineages of mtDNA. Therefore, India is home to three major clades of mtDNA, the M, N and R, which originated in India and populated the rest of the world except Africa (see fig. 17.1).

India therefore possesses the oldest DNA for M macro haplogroup, dated as far back as 73,000–80,000 years before present (ybp), 50,000–70,000 ybp for N and R clades.<sup>16</sup> The M haplogroup shows the greatest variation in India with almost 23 lineages arising directly from the M trunk associated with initial peopling of South Asia during Pleistocene settlement and rules out theories of subsequent colonization events such as the Aryan invasion during the Neolithic Period.<sup>17</sup> The Initial mtDNA pool established in India upon the peopling of India during the Pleistocene has not been replaced, but shaped *in situ* with relatively minor events of gene flow.<sup>18</sup> About 60 per cent of Indians trace their matrilineal heritage in M haplogroup regardless of the caste or tribe groupings. In fact, the M haplogroup is represented by 58 per cent caste, and 72 per cent tribal populations<sup>19</sup> of India, although initially high castes are postulated to show differences.<sup>20</sup>

---

→ Thangaraj et al. 2008, "Maternal Footprints of Southeast Asians in North India", *Human Heredity*, 66(1): 1-9.

Thangaraj et al. 2006; Chaubey et al. 2007: 95; Kivisild et al. 1999a.

<sup>15</sup> Rajkumar et al. 2004; Rajkumar et al. 2005; Kivisild et al. 2004.

<sup>16</sup> Toba volcano in Sumatra erupted 73 kya, covering most of the Indian Ocean region in ash and dark clouds. The genetic evidence, therefore, dates from this period onwards, although archaeological evidence of existence of human settlements is dated much farther back in India. Zielinski et al. 1996, "Potential Atmospheric Impact of the Toba Mega Eruption ~71,000 years ago", *Geophysical Research Letters*, 23: 837-40.

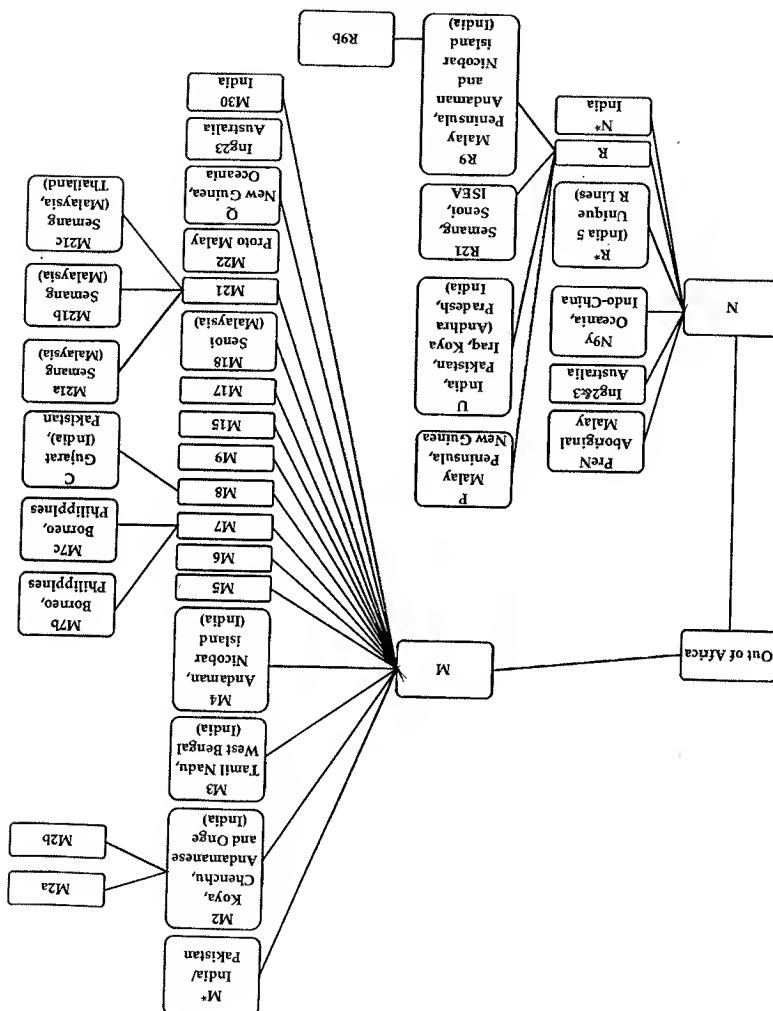
<sup>17</sup> Rajkumar 2004.

<sup>18</sup> Metspalu et al. 2004.

<sup>19</sup> Ibid.

<sup>20</sup> Bamshad et al. 1998.

Fig. 17.1: Female genetic heritage



The haplogroups, M2, M3, M4, M5, M6, M18 and M25 are exclusive to India, as well as all the branches of M haplogroup, M\*, C, D, G, E and Z are observed in Asia. The N haplogroup W is represented by 5 per cent of the population in the north (Gujarat, Punjab and Kashmir). The R haplogroup is represented by its sub-clade U, by its Indian specific sub-branches of haplogroup U2 (U2i: U2a, U2b, U2c throughout India and U7 in Punjab). These U2 clades are represented throughout India at 15 per cent among caste and 8 per cent among tribal populations, which is also represented in Europe.<sup>21</sup> Although the tribal populations of India are classified variously as Negroid, Mongoloid and Australoid based on their physical appearances, they do not show a significant difference in their genetic heritage among the tribal groups, as well as when compared with their neighbours, the caste groups, and their heritage is unrelated to their linguistic heritage.<sup>22</sup> Hence it can be surmised that the difference in physical characteristics among these populations may be the result of lifestyle changes rather than genetic changes mediated by population replacement.

It is impossible to know how the anatomically modern humans looked 80 KYA during their first settlement in India, but it is probable that they may have looked more like the tribal groups rather than the caste groups, since the tribal groups still follow the Upper Palaeolithic lifestyle of the Pleistocene settlers as hunter-gatherers with intermittent swidden agriculture (also known as slash and burn or shifting cultivation), unlike the settled caste groups which changed their lifestyle that may have resulted in physical changes. Since the Andaman and Nicobar islanders remained isolated until the twentieth century, and sustain a prehistoric hunter-gatherer lifestyle, they may closely represent the early settlers, and therefore they have attracted intense scholarly interest in recent efforts, to understand the genetic and demographic features of early settlers of Pleistocene era.

Not much is known about the culture or language or geography of Andaman islands until nineteenth century. However, the term *andam* — from which the name Andamans is derived — resembles *andhram* etymologically, from which is derived the name Andhra Pradesh, the closest coastal state of India for Andaman islands.<sup>23</sup> The first

<sup>21</sup> Kivisild et al. 1999b.

<sup>22</sup> Chaubey et al. 2007; Chaubey et al. 2008.

<sup>23</sup> The term *andhra* is associated with the name of a tribe mentioned in ancient Sanskrit texts. The south-western region of India, through the settlement of this tribe, is named as Andhra region after this tribe, although it is not clear if this designation also included islands off the coast of this region (the closest mainland for Andaman islands is this region). The term *andhaka* and its variant *assaka* were mentioned in Buddhist *jataka* literature. The term *andhra* →

outside intrusion in the Andaman islands was from the penal colony founded by the British regime at Port Blair in 1858 with disastrous consequences for the natives.<sup>24</sup> Of the nineteen linguistic ethnic groups recorded by the British Indian government in 1858 only three groups survived in the nineteenth century, succeeding this intrusion: the Jorawa, Sentinelese and Onge.<sup>25</sup> Colonial anthropologists have noted similarities of Andaman islanders with Pygmies of Africa and Australian aborigines. However, it was not until 1954 that the first clear anthropological study of the Andaman islanders (Onge) was conducted.<sup>26</sup> Scholars suggest however that the Andamanese predate any of the agricultural settlements in this area and hence hold a storehouse of genetic information.<sup>27</sup> The mean Andamanese stature is 1.46 m for males and 1.38 m for females. Their language is related to the languages of east coast of India, particularly Oriya and Telugu. An insular location favoured the survival of this human group and their language, which was not the case with other small-statured groups in South-East Asia, especially, the Aeta of Philippines or the Veddas of Sri Lanka, who are culturally, genetically and linguistically assimilated by their neighbouring populations.

*Genetic Configuration of Indians in  
the Andaman and Nicobar Islands and the Subcontinent*

Two unique mitochondrial DNA haplotypes of M2 and M4 are isolated in the Andamans,<sup>28</sup> with M2 corresponding to the date  $63,000 \pm 6,000$  YBP and M4 corresponding to  $32,000 \pm 7,500$  YBP. So after migrating to the Andaman islands before 63 KYA, the mtDNA of Pleistocene settlers (anatomically modern humans) underwent unique evolution. The M2 haplotypes are also related to the haplotypes from Andhra Pradesh in India.<sup>29</sup> This is remarkable for understanding the spread of human species on earth. The Andaman M4 haplotype also occupies the basal position, in the medial

→ or *andhaka* is used simultaneously to designate a tribe or a region inhabited by this tribe. However, the usage of this term in relation to tribe is notable and coincidentally the tribes of Andhra Pradesh (Chenchu) and the tribes of Andamans hold the most deep-rooted mtDNA lineages of India.

<sup>24</sup> Thangaraj et al. 2003.

<sup>25</sup> Thangaraj 2003: 86; Thangaraj et al. 2006.

<sup>26</sup> Cavalli-Sforza, Menozzi and Piazza 1994: 213.

<sup>27</sup> Oppenheimer 2004; P.A. Underhill, 2003, "Inferring Human History: Clues from Y-chromosome Haplotypes", *Cold Spring Harb. Symp. Quant. Biol.*, 68: 487-93; Underhill et al. 2001.

<sup>28</sup> Endicott et al. 2003a: 181.

<sup>29</sup> Endicott et al. 2006.

lineage joining network, and it is still present in India among the other populations, indicating that it may have been subject to the Pleistocene expansion from the mainland and corresponds with the genetic dating of Indians.<sup>30</sup> While the mtDNA haplogroups specific to Andamans are M31 and M32, the Haplogroup M31 is also found in among the tribal groups of Eastern India (Bihar, Orissa and Andhra Pradesh).

Only M4 haplotypes are reported from Nicobar islands near the Andaman islands.<sup>31</sup> This indicates the occupation of Nicobars may have been coeval with the M4 dates of Andamans (32,000 YBP). The Shompen (population of 200) of Nicobar show a recently derived genetic heritage<sup>32</sup> and their mtDNA reveals two haplogroups of R lineage, B5a and a newly defined clade R12. The Y-chromosome analysis revealed O2a, defined by M95 marker. Thus the genetic make-up of Andamanese and Nicobarese, small-statured isolated human groups of Indian Ocean provides crucial evidence of Pleistocene settlements.

The study outlined above indicates an early Pleistocene settlement in India c. 73,000 YBP and of Andaman islands about 63,000 YBP from the mainland India, on the evidence of the presence of M2 haplotypes in Andaman islands.<sup>33</sup> Later expansion from Andaman islands into Nicobar islands and South-East Asia then onwards can be understood based on the presence of M4 haplotypes in Nicobar islands of India and South-East Asia. Overall the presence of deep-rooted M haplotypes and region-specific haplogroups that indicate *in situ* evolution in Andaman and Nicobar islands indicates the early Pleistocene dispersal of early human groups towards Australia and the role of India as a corridor for such dispersals.

### *Malaysia*

There are two distinct parts to this country, Peninsular Malaysia and East Malaysia. Peninsular Malaysia is located south of Thailand, north of Singapore and east of the Indonesian island of Sumatra. East Malaysia is located on the island of Borneo and shares borders with Brunei and Indonesia. Located near the equator, Malaysia, is characterized by hot and humid climate throughout the year. The Malay Peninsula is located between South-East Asia and North-East Asia. This strategic location is also

<sup>30</sup> Thangaraj et al. 2005; Palanichamy et al. 2006; Thangaraj et al. 2006; Barik et al. 2008.

<sup>31</sup> Prasad et al. 2001.

<sup>32</sup> Trivedi et al. 2006.

<sup>33</sup> Thangaraj et al. 2005; Endicott et al. 2003a, b; Palanichamy et al. 2006, "Phylogeny of Mitochondrial DNA Macrohaplogroup N in India, Based on Complete Sequencing: Implications for the Peopling of Indian Subcontinent", *American Journal of Human Genetics*, 75(6): 966-78.

noted in the genetic trail of the anatomically modern humans, as the first trail out of Malaysia to South-East Asia, while the second trail headed for North-East Asia. Malaysia is unique in two respects: Malaysia preserves an ethno-geographic as well as archaeological continuum from prehistoric times, not noticed in India or Eurasia. In these latter locations the prehistoric archaic humans are replaced by the anatomically modern humans, whereas in Malaysia, the prehistoric humans and Semang share ethno-archaeological and physical characteristics with Neolithic settlers and indicate continuity.<sup>34</sup> The Neolithic hunter-gatherers of Malaysia are considered to be the Semangs<sup>35</sup> based on the cultural, and artefact remains of Perak man. Second, due to its geographical location being centrally located with India on the west, China on the east and the Island South-East Asia (ISEA) in the south, the genetic make-up of Malaysia shows links with all these countries and provides clues to understanding the early dispersal of anatomically modern humans. Therefore, the Semangs emerge as the earliest settlers of this region, as the relict group left from Pleistocene expansion and with respect to genetics, archaeology and physical characteristics.

Malaysia is a multiethnic society with several groups clumped together as Malays, Indians, Chinese and aborigines. There are almost twenty aboriginal groups known as *Orang Asli*<sup>36</sup> forming 5 per cent of the Malaysian population.<sup>37</sup> The *Orang Asli* (original humans) is further divided into three main categories as Semang, Senoi and proto-Malays. Negrito was a recent term employed in describing the populations previously known as Semang (Perak and Pahang) and Pangan (Kelantan). These terms, although come to designate tribes, were not originally tribal names or racial designations. Pangan is also sometimes employed to refer to the aboriginal Malays. The word *pangan*, means "forest people", while *semang*, means "debt slaves" or dependents.<sup>38</sup> Semang may also mean "human being" a derivative of Central Aslian language "Sema". However, the north-western Semang groups refer to themselves as *meni* while the south-eastern groups refer to themselves as *batek*, meaning "human beings". Semangs speak several dialects of the Aslian language, related to the Mon-Khmer group of languages.

Semang are noted for their resemblance to Africans and Australians and other

<sup>34</sup> Hill et al. 2006; Oota et al. 2001. Oota et al. establish that there is a direct continuity between pre-Neolithic humans and the present Semang and that the Neolithic humans in this area might be an ancestral group of the Senoi.

<sup>35</sup> Bulbeck, David 2005; Majid and Tjia 1988; Matsumura, Hirofumi and Majid 1995.

<sup>36</sup> Carey 1976; Benjamin 2002.

<sup>37</sup> Cary 1976: 101, 126.

<sup>38</sup> Ibid.: 5.

aboriginal groups of South-East Asia such as the Aeta of Philippines and men average in height about 153 cm (5 ft 1/2 in) while women average around 142.7 cm (4 ft 8 in) in height.<sup>39</sup> The Semang and Pangan are nomadic while Senoi are swidden agriculturalists and the Senoi are classified as Australoid or Veddoid.<sup>40</sup> The proto-Malays are Mongoloid and show a recent northeast Asian admixture, and are termed Bhūmiputras and classified as Mongoloid tribe. The population of Semang was 3,200, while that of Senoi was 49,000, and that of proto-Malays was 42,000 in 2000.<sup>41</sup> These three groups, although classified as belonging to different racial groups, as Negroid, Australoid or Mongoloid, show similar genetic lines derived from M and N clades.

#### *Genetic Configuration of the Malaysian Orang Asli*

Three quarters of the Semang consist of a unique, mtDNA, M, and N lines that relate them to the African exodus and early beach settlers in India, especially the Andaman islands.<sup>42</sup> Their N line relates to the R1 mitochondrial DNA evolved in India. This unique genetic make-up is not known in any other East Asian or South-East Asian populations, which links it to its westerly origin in India and places it at the basal position. It can be deduced that after arriving from India, these populations may have existed in Malaysia in total isolation from any other neighbouring population during the Pleistocene era and until the modern day that, they do not share their DNA mark up with other East Asians. The other aborigines of Malaysia known as proto-Malays share their DNA make-up with other East Asian populations, which indicate recent migration, about the Neolithic Period, 3 or 4 KYA.

Investigations of mtDNA of Malaysia reveal Pleistocene settlement in Malaysia close to 60,000 YBP.<sup>43</sup> The *Orang Asli* of Malaysia show M, N and R haplogroups (M, N and R originated in India). While the Mendrik and Batek Orang Asli show M21a, the Jahai Semang show R21 clade. Within the M haplogroup, there is an ancient and highly localized clade, M21, with three derived sub-clades (A, B, C) is dated to 57,000 YBP.<sup>44</sup> M21a is the most common in the Semang (almost 84 per cent in the Mendrik Semang) and it is also present in the "Maniq" Semang (*mani* or *asli* of Thailand) from southern

<sup>39</sup> Cary 1976: 14-15.

<sup>40</sup> Ibid.: 1976: 20.

<sup>41</sup> Benjamin 2002.

<sup>42</sup> Oppenheimer 2004: 169

<sup>43</sup> Macaulay 2005: 1034-36.

<sup>44</sup> Macaulay et al. 2005.

Thailand suggesting a colonization event from Malaysia. The Thai Semang also share genetic make-up of M21a, and it is also found in minority in the Batek and Jahai Semang of Malaysia. The M21b shares a common ancestor with M21a and dates to 44,000 YBP. It is shared between Semang and Senoi and is also noticed among proto-Malay aborigines and Island South-East Asia. This may indicate a long-standing relationship among distinct aboriginal groups. M22 haplogroup is found in Temuan Semang and also noticed Thailand. Haplogroup R21 (derivative of N, which originated in northwest India some time after emigrating from Africa) is present only in the Semang (mainly Jahai). R21 is not noticed in any other groups. Therefore R21, like M21, is indigenous to Semang/Senoi. Being the oldest occupants and the least populous group of the region the Semang have lost most of their genetic diversity through genetic drift and genetic selection. Although the Semang represents the Upper Pleistocene ancestry in the Malay Peninsula, their M and N lines failed to disperse more widely due to the small population size.

The Semang appear to be the most direct descendants of the original inhabitants of the Malay peninsula and have experienced only minor genetic admixture. Like the Andaman islanders and other tribal groups of India who evolved *in situ*, the Malayan Orang Asli groups also indicate internal development after the initial expansion during the Pleistocene from the first migration of anatomically modern humans.

#### *Island South-East Asia (ISEA)*

Island South-East Asia (ISEA) which is encompassed by Indonesia, East Malaysia and the Philippines, was first reached and settled by anatomically modern humans about 45,000 YBP. At that time the region was split between the Pleistocene continent of Sunda, which stretched from Sumatra to Bali and Palawan, and Wallacea, along with the islands east of Wallace's line.<sup>45</sup>

#### *Genetic Configuration of ISEA*

ISEA contains its own M haplogroups, derived from the initial settlement during the Pleistocene era. The most commonly found and indigenous haplogroup E is derived from M and dates back to 25,000 YBP with sub-clades ranging from 6000–16,000 years ago.<sup>46</sup> Another subgroup of M haplogroup found in ISEA, haplogroup Q, occurs with

<sup>45</sup> M. Bird, D. Taylor and C. Hunt 2005, "Paleo-environments of Insular Southeast Asia During the Last Glacial Period: A Savanna Corridor in Sundaland?", *Quaternary Science Review*, 24: 2228–42.

<sup>46</sup> Soares et al. 2008.

three sub-clades which indicate deep antiquity in the region.<sup>47</sup> This haplogroup is found predominantly in New Guinea and in low levels in remote Oceania.<sup>48</sup> At least three basal M groups, M45, 46, 47 are identified in ISEA, plus M21d, a novel branch of M21 (found commonly in Malaysia) is identified in ISEA. This is also found in high frequency in Austronesian-speaking Moken "Sea gypsies" of Burma. All these M haplogroups are very ancient and are dated between 40,000–58,000 YBP. Almost 14 per cent the mtDNA of the ISEA contains M haplogroups indicating Pleistocene colonization of these islands by anatomically modern humans.<sup>49</sup> Haplogroup N21 indicates a closer connection with Malaysian Semang and R22 indicates a connection with Shompen group of Nicobar islands of India, although it is most diverse in ISEA. The mtDNA haplogroups M45, 46, 47, N21d and R22 are indigenous to ISEA,<sup>50</sup> they also indicate the dispersal of anatomically modern humans across the Indian Ocean and support the theory that the Andamanese and Semang of Malaysia and Thailand represent the relict groups of early human migrations from India throughout the Indian Ocean region.

The female genetic heritage is clear and represents *in situ* origin of Indian lines, which later spread across the Indian Ocean region.<sup>51</sup> However, the male genetic heritage is not as clear and hence utilized by earlier scholars in support of the Aryan migration/invasion theories. I discuss below the male genetic heritage of India and other Indian Ocean regions in order to examine the genetic heritage.

#### ARYAN MIGRATION THEORIES, MALE GENETIC HERITAGE OF INDIA, AND INDIAN OCEAN REGION

Initial findings of genetic research are utilized to support the existing theories of Aryan invasion/migration into India. However, these theories lost ground as more research on genetic heritage of India became available, and the Y-chromosome heritage mostly substantiated the findings of population migrations discussed above. I review the basic corpus of such earlier scholarship here.

<sup>47</sup> Thangaraj et al. 2008.

<sup>48</sup> Hill et al. 2006.

<sup>49</sup> Hill et al. 2007.

<sup>50</sup> Hudjashov et al. 2007.

<sup>51</sup> Kivisild 1999a, b.

*The Question of Indo-Aryans and Presence/Absence of Their Genetic Heritage in India*

For reasons that are unknown and unexplained a group of scholars tend to assume that the Indo-Aryan migrations, occurred about 30-40 KYA, and are male driven.<sup>52</sup> Hence the initial research focused on finding common male genetic heritage lines between India, Caucasus and Europe, the supposed regions where the Aryan tribes were said to have spread. One such male genetic line is M17 (R1a), one of the most frequent genetic lines found in India and Pakistan, but also across the Caucasus and Europe. M17 is found widely in Eastern Europe, its highest frequency is noticed in Hungary (60 per cent) while it is also found in low frequencies in western and southern Europe. M17 is also found in India in high frequencies including southern India. Accordingly, it was proposed that M17 arose in Altai region giving rise to the European Y lines, which later found its way to India, thus authenticating the arrival of Indo-Aryans. In order for this theory to work, two things are necessary: First, M17 should originate somewhere outside the borders of India, and second, the genetic lines outside India should show more diversity and deep-rooted, and older genetic heritage. First point finds no evidence outside India, since the origin of M17 was not located outside of India, but placed inside of India.<sup>53</sup> Although M17 was initially thought of as a marker of "male Aryan invasion", of India, its origin in India proves the contrary. Second, that the frequency and deep-rooted heritage is accounted for M17 in India through subsequent research, which also supports its original home as India.<sup>54</sup>

High frequencies of M17/R1a among the Punjabis of north-west India and Chenchus of Andhra Pradesh in south-eastern India indicate an ancient and shared genetic heritage of the subcontinent. M17 is dated to 36 KYA in India, while the oldest evidence of its occurrence in Europe could only be dated to 23 KYA. This only makes the possible spread of M17 from India rather than vice versa as thought of earlier. The M17 could have found its way initially from India to Pakistan via Kashmir and then westward on to Europe via the Caucasus. M17 (R1a) as a marker of Indo-Aryan migration<sup>55</sup> is not viable in the light of the above genetic evidence.<sup>56</sup> These R1a migrations originate in

<sup>52</sup> Bamshad et al. 2001; R. Cordaux et al. 2004, "Independent Origins of Indian Caste and Tribal Paternal Lineages", *Current Biology*, 14: 231-35.

<sup>53</sup> Kivisild et al. 2003.

<sup>54</sup> Sengupta et al. 2006.

<sup>55</sup> Well et al. 2001.

<sup>56</sup> Underhill et al. 2010.

the north-western India, and appear in Caucasus up to Siberia, and in Eastern Europe up to Hungary and Poland, a wide geographical region with varying frequencies ranging between 27-56 per cent of the population.<sup>57</sup> In addition to this ancient gene flow dated to Holocene period a recent migration from India during the current millennium is also noticed in Eastern Europe and Caucasus regions among Roma populations, which share close genetic affinity with India.<sup>58</sup> In addition, M17 is found in high frequencies in India, including south India, among the caste and tribal groups showing a horizontal distribution, which disproves any high or low-caste distinctions.<sup>59</sup> While its occurrence among the Chenchus of Andhra Pradesh is ruled out as aberration, it cannot be ruled out or ignored so easily as it also shows wide-ranging diversity indicating its ancient status in numerous other tribal and caste groups in eastern India.<sup>60</sup> The deep-rooted ancestry of tribes and their genetic heritage at the basal level are discussed above in part I of this paper, hence I will confine my discussion to the origins and spread of M17 here. While the European M17 is less diverse and dated only from 23,000 ybp, it is dated to 36,000 ybp indicates its deep-rooted nature in India. This early date also corresponds to the Microlithic phase of human settlements in India noted for demographic and settlement increase in India.<sup>61</sup> All this suggests that the M17 originated somewhere in south India (especially among the tribals — who can be considered the relict populations) and found its way to north-west India through Kashmir, spreading from there to Central Asia and then to Europe.

Consensus is noted among the historians and archaeologists of India that the Pleistocene migrations of Anatomically Modern Humans replaced the Archaic Humans of Lower Palaeolithic period by Upper Palaeolithic period and Mesolithic India is completely populated by Anatomically Modern Humans.<sup>62</sup> Gene pool of Indians could be traced to this early migrations and settlements dated between 56,000–36,000 ybp. Y-chromosomal haplogroups of this initial settlement C, F, K and D, and several sub-clades of F and K such as H, L, R2 and F are commonly found in India. No gene pool that is specific to upper castes at about 50 per cent of the population datable to 4 or 5 KYA is found in India.

<sup>57</sup> Primorac et al. 2011.

<sup>58</sup> Gresham et al. 2001.

<sup>59</sup> Kivisild et al. 2003.

<sup>60</sup> Cordaux et al. 2003.

<sup>61</sup> Petraglia et al. 2009.

<sup>62</sup> Joshi 1994; Petraglia et al. 2009.

To support the origin of M17 in India one also need to look at its ancestor line M173 (R) from which the M17 originated and also P haplogroup from which M173 originated. M173 is spread throughout the old world from England to South and Central Asia, Siberia, Australia and even to America. This is the most common single male genetic line in the world. M173 along with its branch M17 originated in India which can be proved based on its diversity and deep-rooted ancestry in India, it not only migrated north splitting up to spread into Europe, but also to East Asia and Australia (Semino et al. 2000).<sup>63</sup> The male genetic line of M173 also has very ancient origin in India and spread from India to Europe 30 KYA by a similar route as the M17 noted above. Unlike M17, its ancestor M173 is a common genetic marker in Europe reaching frequencies of up to 86 per cent in the Basques and British Isles. Therefore, M173 represents the original upper Palaeolithic invasions of Anatomically Modern Humans from East to West, and not vice versa.

Both of these lines, M17 and M173, account for 50 per cent of the extant male genetic lines in Europe, and are spread from India during the Upper Palaeolithic period around 33-30 KYA. Therefore the male originated theory of genetic heritage of Aryan invasion/migration theories is disproved.

Another argument in support of Aryan migrations was that the castes and tribes have different genetic heritage, and that the many castes of India represent the various waves of Aryan migration that took place over a long period of time. In support of this theory, it was proposed that some haplogroups (J2, R1a, R2 and L) are associated with the migration of Indo-Aryans into India. It is assumed that "the upper castes are more similar to Europeans than to Asian, the middle castes are equidistant from the two groups, and lower castes are most similar to Asians".<sup>64</sup> This argument also does not stand reasoning in the light of the genetic heritage discussed above. The Haplogroups R1a (M17), R2 and L are also derivatives of deep-rooted M173 (R), which had its origin in India, these haplogroups are genetically deep-rooted in India than Europe. These haplogroups (R1a, R2 and L) also occur among the caste as well as tribal population with similar frequencies, while diversity is greater among the tribal groups, which supports the view that the gene flow is from tribes to castes is common, and a common phenomenon but not isolated demographic feature.<sup>65</sup> And hence the castes may not

<sup>63</sup> Semino et al. 2000.

<sup>64</sup> Bamshad et al. 2001: 996.

<sup>65</sup> Sharma et al. 2012, "Genetic Affinities of the Central Indian Tribal Populations", *Plos One*, 7(2).

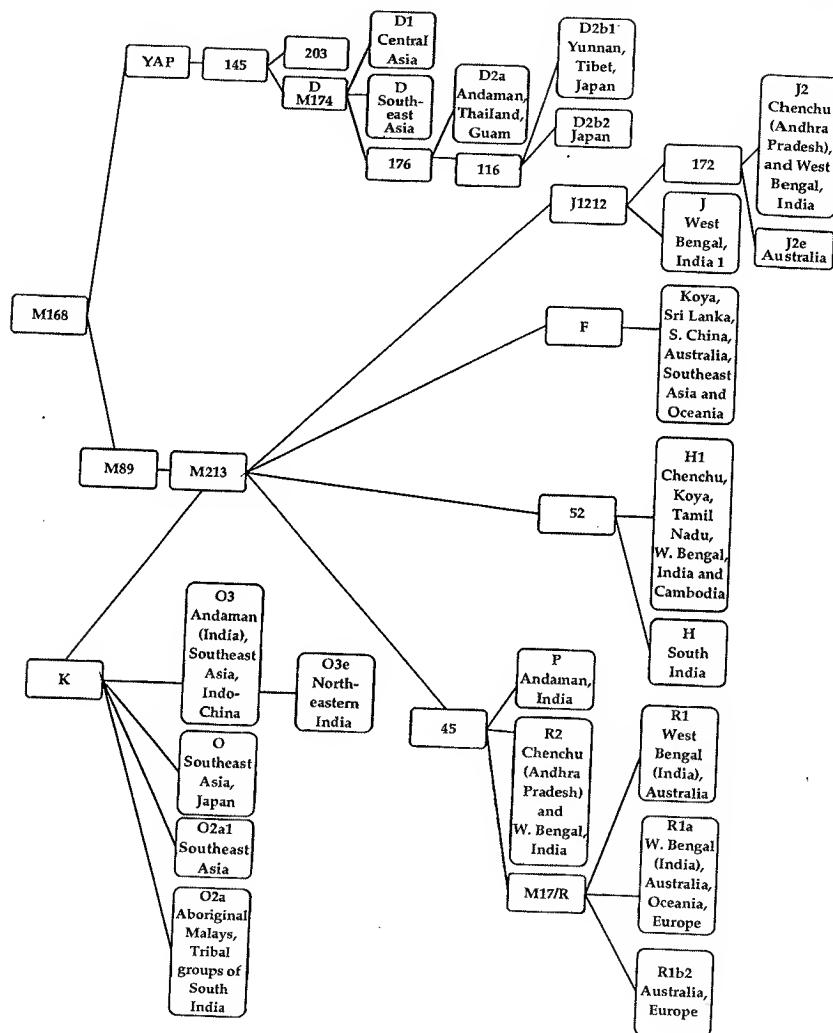


fig. 17.2: Male genetic heritage

have originated independently of the tribes in India. Both caste and tribes have similar origins and spread, and in fact tribes hold the most deep-rooted ancestor lines, which disproves any migration, but only coexistence and evolution of castes and tribes within India.<sup>66</sup> The caste designations and languages used by various groups in India are only sociological and cultural adaptations, but do not indicate any genetic subdivisions.

This then brings one to the next question: does South Asia contain the founder genetic heritage of all non-African genetic lines? The existence of short-statured human groups in the Indian Ocean region and their deep-rooted genetic heritage might prove in affirmative for the above question. I will examine the male genetic heritage of the short statured human groups below to understand the initial peopling of ancient human groups in India as well as the world.

#### *Male/Y-Chromosome Lineages in the Indian Ocean Region*

Similar to the mtDNA evidence (X-chromosome inherited from the mother) discussed above, the Y-chromosome (inherited from the father) also provides unique evidence regarding the early origin and spread of anatomically modern humans in the Indian Ocean region (see fig. 17.2).<sup>67</sup> M130 arose in Asia some time after the early departure from Africa.<sup>68</sup> The M130 distribution in the Indian Ocean rim is consistent with the theory of coastal route for the migration of anatomically modern humans eventually reaching Australia.<sup>69</sup> India carries major M130 lineages,<sup>70</sup> which provides indirect evidence for a coastal route via South-East Asia and Oceania up to Australia.<sup>71</sup> Several studies have also shown the presence of M130 derived chromosomes in Melanesia, Australia and Polynesia.<sup>72</sup>

Based upon phylogenetic geography, it has been postulated that Y-Alu Polymorphism (YAP) arose in Africa from M168 or an ancestor and later these YAP departed from Africa in an early dispersal event<sup>73</sup> and evolved into M174 persistent in

<sup>66</sup> Sahoo et al. 2006, "A Prehistory of Indian Y-chromosomes: Evaluating Demic Diffusion Scenarios", *Proceedings of the National Academy of Sciences of the United States of America* 103(4): 843-48.

<sup>67</sup> Cordaux et al. 2003; Karafet 2008.

<sup>68</sup> Underhill et al. 2001.

<sup>69</sup> Cordaux et al. 2003; Oppenheimer 2005.

<sup>70</sup> Kivisild et al. 2003; Redd et al. 2002.

<sup>71</sup> Birdsell 1993.

<sup>72</sup> Karafet et al. 1999; Capelli et al. 2001.

<sup>73</sup> Underhill et al. 2001.

Asia.<sup>74</sup> Recent molecular analysis of four Jorawa and twenty-four Onge males from Andaman Islands revealed that they all belonged to the M174 defined D haplogroup.<sup>75</sup> The presence of distinctive M174 lineages in the Andaman islands, Japan and Asian mainland indicates that the Andamanese populations have been isolated geographically for a considerable time and played a key role in the Pleistocene migrations. The M174/D haplotype data supports the coastal migration model of the dispersal of anatomically modern humans from Africa during Pleistocene episodes taking advantage of the lower sea level<sup>76</sup> and also places the small statured humans of India, especially the Onge, in a key position in assessing the migration and settlement routes of anatomically modern humans.

Both M130 and M174-related lineages have different frequencies and geographic distribution patterns and are reflective of the early formation of non-African heritage. In this pattern the Andaman islands supply the critical evidence necessary to understand the coastal migration of early humans out of India towards Australia, and northwards towards the rest of the world placing India in a basal position.

#### MT DNA AND Y-CHROMOSOME LINEAGES: KEY IMPLICATIONS FOR INDIA

Mitochondrial DNA and Y-chromosome analysis place India in a key position for understanding the origin of all non-African populations and provides evidence for the late Pleistocene human migrations reaching the West by the Holocene period. All the three clades of L3 M and N, and R are a derivative of N in India indicates that India is in the basal position for Pleistocene population migrations. Superhaplogroup N branch is commonly found in East Asia, through its sub-clade R, which later differentiated into two clusters in East Asia.<sup>77</sup> Indian M groups are considerably older and more diverse than South-East Asian or Australian haplogroups. In a similar way, Y-chromosomal haplogroups C, F, K and D (in Andaman islands) and several sub-clusters of F and K (H, L, R2 and F), also indicate the Pleistocene migration across India and Indian Ocean region. Australian aborigines and Indians share 50 per cent C chromosome common ancestry and C is also ubiquitous in Indo-Pacific populations which also indicates the Pleistocene migration from India via Island corridor reaching Australia during the Holocene.<sup>78</sup> After this initial settlement process, each region

<sup>74</sup> Underhill 2003.

<sup>75</sup> Thangaraj et al. 2003.

<sup>76</sup> Underhill et al. 2003; Oppenheimer 2005.

<sup>77</sup> Kivisild et al. 1999a.

<sup>78</sup> Redd et al. 2002; Hudjashov et al. 2007; Kivisild 1999b.

(including India) developed its region-specific branches of these founders<sup>79</sup> and continued to replicate and evolve *in situ* and genetic evidence does not support any great migrations or invasions after this early Pleistocene settlement in India.

The small-statured human groups discussed above share more physical resemblances and are connected by basal genetic lineages, although with specific individual genetic make-up. But this is a result of progressive genetic change natural for the settled human groups in any geographical area. It should be understood that, it is their Pleistocene ancestors that are related, but not the modern humans in various location, they are only the relict groups of the early migrations. Therefore, even though the Anatomically Modern Humans are represented by the short-statured human groups as they progressed from one location to another on the path of their migration across the globe, their genetic make-up evolved uniquely *in situ* in each region, while retaining some of the original physical characteristics that connect them all. In this basal position holding the most ancient genetic signature in a number of locations, the small-statured human groups remain to this day and can be considered the ancestral representatives of all non-African humanity. Therefore, India and the rest of the world are populated by similarly short-statured human groups, with no further migrations of any human groups (30-40 KYA), let alone the tall-statured Aryans groups as proposed earlier. Later physical features of various Indians only reveal adaptation, lifestyle changes and evolution in each region.

The size of the human beings increases the farther they are located from the nuclear area of these first major steps taken by the Pleistocene humans, evidence of which is preserved in the form of the small-statured humans groups present in the Indian Ocean region discussed in this paper. It is then clear that all the non-Africans evolved from the small-statured human groups, spread from India, but not vice versa.

#### Acknowledgement

First of all I would like to thank Nalini Rao for inviting me to contribute to this volume. I would like to thank Edwin Bryant for reading an earlier version of this paper and for his valuable advice. I would like to thank my friend R.L. Mohl, Interpretive Specialist, Hopewell Mound Historical Site, Hopewell, Ohio, for reading an earlier version of this paper. Finally I would like to thank my father Mandalapu Vijaya Bhaskara Rao for teaching me to think differently and for his many discussions on the Aryans.

<sup>79</sup> Kivisild 1999b.

This paper is the result of a historian's attempt to utilize published research data on human genetics to understand the implications for the early history of India, especially, the question of Aryans that has plagued the history of India for more than two centuries. I thank all the authors of the genetic research papers utilized in this paper for their meticulous research and publications. However, any errors in this paper are mine.

## References

Bamshad, M.J., W.S. Watkins, M.E. Dixon, L.B. Jorde, B.B. Rao, J.M. Naidu, B.V.R. Prasad, A. Rasanayagam and M.F. Hammer, 1998, "Female Gene Flow Stratifies Hindu Castes", *Nature*, 395(6703): 651-52.

Bamshad, M., T. Kivisild, W.S. Watkins, M.E. Dixon, C.E. Ricker, B.B. Rao, J.M. Naidu, B.V.R. Prasad, P.G. Reddy, A. Rasanayagam, S.S. Papiha, R. Villem, A.J. Redd, M.F. Hammer, S.V. Nguyen, M.L. Carroll, M.A. Batzer and L.B. Jorde, 2001, "Genetic Evidence on the Origins of Indian Caste Populations", *Genome Research*, 11(6): 994-1004.

Barik, S.S., R. Sahani, B.V.R. Prasad, P. Endicott, M. Metspalu, B.N. Sarkar, S. Bhattacharya, P.C.H. Annapoorna, J. Sreenath, D. Sun, J.J. Sanchez, S.Y.W. Ho, A. Chandrasekar and V.R. Rao, 2008, "Detailed mtDNA Genotypes Permit Reassessment of the Settlement and Population Structure of the Andaman Islands", *American Journal of Physical Anthropology* 136(1): 19-27.

Behar, Doron M., Mannis van Oven, Saharon Rosset, Malt Metspalu, Eva-LiisLoogvali, Nuno M. Silva, Toomas Kivisild, Antonio Torroni and Richard Villem, 2012, "A 'Copernican' Reassessment of the Human Mitochondrial DNA Tree from Its Root", *American Journal of Human Genetics*, 90(4): 675-84.

Benjamin, G., 2002, "On Being Tribal in Malay World", *Tribal Communities in the World: Historical, Cultural and Social Perspectives*, ed. G. Benjamin and C. Chou, pp. 7-76, Singapore: Institute of Southeast Asian Studies.

Bird, M., D. Taylor and C. Hunt, 2005, "Paleo-environments of Insular Southeast Asia During the Last Glacial Period: A Savanna Corridor in Sundaland?", *Quaternary Science Review*, 24: 2228-42.

Birdsell, J.B., 1993, *Microevolutionary Patterns in Aboriginal Australia: A Gradient Analysis of Clines*, Oxford: Oxford University Press.

Bryant, E., 2001, *The Quest for Origins of Vedic Culture and the Indo-Aryan Migration Debate*, New York: Oxford University Press.

Bryant, Edwin F. and Laurie L. Patton, 2005, *The Indo-Aryan Controversy: Evidence and Inference in Indian History*, London: Routledge.

Bulbeck, David, 2005, "The Gua Cha Burials", in *The Perak Man and Other Prehistoric Skeletons of Malaysia*, pp. 253-309, Penang (Malaysia): Penebit University Sains Malaysia.

Bulbeck, D., P. Raghavan and D. Rayner, 2006, "Races of Homo Sapiens: If not in the Southwest Pacific, then Nowhere", *World Archaeology*, 38(1): 109-32.

Bulbeck, David, 2007, "Where River Meets Sea: A Parsimonious Model for Homo sapiens Colonization of the Indian Ocean Rim and Sahul", *Current Anthropology*, 48(2): 315-21.

Capelli, C., J.F. Wilson, M. Richards, M.P.H. Stumpf, F. Gratrix, S. Oppenheimer, P. Underhill, V.L. Pascali, T.M. Ko and D.B. Goldstein, 2001, "A Predominantly Indigenous Paternal Heritage for the Austronesian-speaking Peoples of Insular Southeast Asia and Oceania", *American Journal of Human Genetics*, 68(2): 432-43.

Carey, Iskandar, 1976, *Orang Asli: The Aboriginal Tribes of Peninsula Malaysia*, Kaula Lumpur (Malaysia): Oxford University Press.

Cavalli-Sforza, L., P. Menozzi and A. Piazza, 1994, *The History and Geography of Human Genes*, Princeton, NJ: Princeton University Press.

Chaubey, Gyaneshwer, Monika Karmin, Ene Metspalu, Mait Metspalu, Deepa Selvi-Rani, Vijay Kumar Singh, Jueri Parik, Anu Solnik, Prathap Naidu, Ajay Kumar, Niharika Adarsh, Chandana Basu Mallick, Bhargav Trivedi, Swami Prakash, Ramesh Reddy, Parul Shukla, Sanjana Bhagat, Swati Verma, Samiksha Vasnik, Imran Khan, Anshu Barwa, Dipti Sahoo, Archana Sharma, Mamoon Rashid, Vishal Chandra, Alla G. Reddy, Antonio Torroni, Robert A. Foley, Kumarasamy Thangaraj, Lalji Singh, Toomas Kivisild and Richard Villems, 2008a, "Phylogeography of mtDNA haplogroup R7 in the Indian Peninsula", *BMC Evolutionary Biology*, 8: 227.

Chaubey, Gyaneshwer, Mait Metspalu, Monika Karmin, Kurnarasarny Thangaraj, Siiri Roots, Juri Parik, Anu Solnik, Deepa Selvi-Rani, Vijay Kumar Singh, B. Prathap Naidu, Alla G. Reddy, Ene Metspalu, Lalji Singh, Toomas Kivisild and Richard Villems, 2008b, "Language Shift by Indigenous Population: A Model Genetic Study in South Asia", *International Journal of Human Genetics*, 8(1-2): 41-50.

Chaubey, Gyaneshwer, Mait Metspalu, Ying Choi, Reedik Maegi, Irene Gallego Romero, Pedro Soares, Mannis van Oven, Doron M. Behar, Siiri Roots, Georgi Hudjashov, Chandana Basu Mallick, Monika Karmin, Mari Nelis, Jueri Parik, Alla Goverdhana Reddy, Ene Metspalu, George van Driem, Yali Xue, Chris Tyler-Smith, Kumarasamy Thangaraj, Lalji Singh, Maiti Remm, Martin B. Richards, Marta Mirazon Lahr, Manfred Kayser, Richard Villems and Toomas Kivisild, 2011, "Population Genetic Structure in Indian Austroasiatic Speakers: The Role of Landscape Barriers and Sex-Specific Admixture", *Molecular Biology and Evolution*, 28(2): 1013-24.

Chaubey, Gyaneshwer, Mait Metspalu, Toomas Kivisild and Richard Villems, 2007, "Peopling of South Asia: Investigating the Caste-tribe Continuum in India", *Bioessays*, 29(1): 91-100.

Cordaux, R. and M. Stoneking, 2003, "South Asia, the Andamanese, and the Genetic Evidence for an "Early" Human Dispersal Out of Africa", *American Journal of Human Genetics*, 72(6): 1586-90.

Cordaux, R., Robert Aunger, Gillian Bentley, Ivane Nasidze, S.M. Sirajuddin and Mark Stoneking, 2004, "Independent Origins of Indian Caste and Tribal Paternal Lineages", *Current Biology*, 14: 231-35.

Datta-Roy, A. and Praveen Karanth, 2009, "The Out of India Hypothesis: What Do Molecules Suggest?", *Journal of Biosciences*, 34(5): 687-97.

Endicott, P., M.T.P. Gilbert, C. Stringer, C. Lalueza-Fox, E. Willerslev, A.J. Hansen and A. Cooper,

2003a, "The Genetic Origins of the Andaman Islanders", *American Journal of Human Genetics*, 72(1): 178-84.

Endicott, P., V. Macaulay, T. Kivisild, C. Stringer and A. Cooper, 2003b, "South Asia, the Andamanese, and the Genetic Evidence for an 'Early' Human Dispersal Out of Africa — Reply", *American Journal of Human Genetics*, 72(6): 1590-93.

Endicott, Phillip, Mait Metspalu, Chris Stringer, Vincent Macaulay, Alan Cooper and Juan J. Sanchez, 2006, "Multiplexed SNP Typing of Ancient DNA Clarifies the Origin of Andaman mtDNA Haplogroups Amongst South Asian Tribal Populations", *Plos One*, 1(1): 281.

Endicott, Phillip, Mait Metspalu and Toomas Kivisild, 2007, "Genetic Evidence on Modern Human Dispersals in South Asia: Y-chromosome and Mitochondrial DNA Perspectives: The World Through the Eyes of Two Haploid Genomes", in *Evolution and History of Human Populations in South Asia*, ed. M.D. A.B. Petraglia, pp. 229-44, Dordrecht, Netherlands: Springer.

Gresham, D., B. Morar, P.A. Underhill, G. Passarino, A.A. Lin, C. Wise, D. Angelicheva, F. Calafell, P.J. Oefner, P. Shen, I. Tournev, R. de Pablo, V. Kučinskas, A. Perez-Lezaun, E. Marushiaikova, V. Popov and L. Kalaydjieva, 2001, "Origins and Divergence of the Roma (Gypsies)", *American Journal of Human Genetics*, 69(6): 1314-31.

Hill, Catherine, Pedro Soares, Maru Mormina, Vincent Macaulay, Dougie Clarke, Petya B. Blumbach, Matthieu Vizuet-Forster, Peter Forster, David Bulbeck, Stephen Oppenheimer and Martin Richards, 2007, "A Mitochondrial Stratigraphy for Island Southeast Asia", *American Journal of Human Genetics*, 80(1): 29-43.

Hill, Catherine, Pedro Soares, Maru Mormina, Vincent Macaulay, William Meehan, James Blackburn, Douglas Clarke, Joseph Maripa Raja, Patimah Ismail, David Bulbeck, Stephen Oppenheimer and Martin Richards, 2006, "Phylogeography and Ethnogenesis of Aboriginal Southeast Asians", *Molecular Biology and Evolution*, 23(12): 2480-91.

Hudjashov, G., T. Kivisild, P.A. Underhill, P. Endicott, J.J. Sanchez, A.A. Lin, P. Shen, P. Oefner, C. Renfrew, R. Villemans and P. Forster, 2007, "Revealing the Prehistoric Settlement of Australia by Y-chromosome and mtDNA Analysis", *Proc. Natl. Acad. Sci. USA*, 104(21): 8726-30.

*India Yearbook 2007*, Government of India, Publications Division, New Delhi: Ministry of Information and Broadcasting.

Joshi, R., 1994, "South Asia in the Period of *Homo Sapiens Neandertalensis*; South Asia in the Period of *Homo Sapiens Sapiens*", *History of Humanity*, vol. 1: *Prehistory and the Beginnings of Civilization*, ed. De Laet, Paris, London: Unesco and Routledge, pp. 162-65; 256-59.

Karafet, T.M., S.L. Zegura, O. Posukh, L. Osipova, A. Bergen, J. Long, D. Goldman, W. Klitz, S. Harihara, P. de Kniff and V. Wiebe, 1999, "Ancestral Asian Sources of New World Y-Chromosome Founder Haplotypes", *American Journal of Human Genetics*, 64: 817-31.

Karafet, T.M., F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura and M.F. Hammer, 2008, "New Binary Polymorphisms Reshape and Increase Resolution of the Human Y-chromosomal Haplogroup Tree", *Genome Res.*, 18(5): 830-8.

Khaliq, S. et al, 2007, "Y-chromosomal Evidence for a limited Greek Contribution to the Pathan Population of Pakistan", *Eur.J.Hum.Genet.*, 15(1): 121-26.

Kivisild, T., M.J. Bamshad, K. Kaldma, M. Metspalu, E. Metspalu, M. Reidla, S. Laos, J. Parik, W. S. Watkins, M.E. Dixon, S.S. Papiha, S.S. Mastana, M.R. Mir, V. Ferak and R. Villems, 1999a, "Deep Common Ancestry of Indian and Western-Eurasian Mitochondrial DNA Lineages", *Current Biology*, 9(22): 1331-34.

Kivisild, T., K. Kaldma, M. Metspalu, J. Parik, S. Papiha and R. Villems, 1999b, "The Place of the Indian Mitochondrial DNA Variants in the Global Network of Maternal Lineages and the Peopling of the Old World", in Surinder S. Papiha, Ranjan Deka and Ranajit Chakraborty (eds.) *Genomic Diversity: Applications in Human Population Genetics*, New York: Kluwer Academic/Plenum.

Kivisild, T., M. Reidla, E. Metspalu, A. Rosa, A. Brehm, E. Pennarun, J. Parik, T. Geberhiwot, E. Usanga and R. Villems, 2004, "Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears", *American Journal of Human Genetics*, 75(5): 752-70.

Kivisild, T., S. Roots, M. Metspalu, S. Mastana, K. Kaldma, J. Parik, E. Metspalu, M. Adojaan, H. V. Tolk, V. Stepanov, M. Golge, E. Usanga, S.S. Papiha, C. Cinnioglu, R. King, L. Cavalli-Sforza, P.A. Underhill and R. Villems, 2003, "The Genetic Heritage of the Earliest Settlers Persists both in Indian Tribal and Caste Populations", *American Journal of Human Genetics*, 72(2): 313-32.

Lee, S.S., J. Mountain, B. Koenig, R. Altman, M. Brown, A. Camarillo, L. Cavalli-Sforza, M. Cho, J. Eberhardt, M. Feldman, R. Ford, H. Greely, R. King, H. Markus, D. Satz., M. Snipp, C. Steele and P. Underhill, 2008, "The Ethics of Characterizing Difference: Guiding Principles on Using Racial Categories in Human Genetics", *Genome Biol.*, 9(7): 404.

Macaulay, Vincent, Catherine Hill and A. Achilli, 2005, "Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes", *Science*, 308: 1034-36.

Majid, Z. and H.D. Tjia, 1988, "Kota Tampan, Perak: The Geological and Archaeological Evidence for a Late Pleistocene Site", *Journal of the Malaysian Branch of the Royal Asiatic Society*, 61: 123-34.

Matsumura, Hirofumi and Zuraina Majid, 1995, "Metrical Analysis of the Dentition of Perak Man from Gua Gunung Runtuh in Malaysia", *Bulletin of the National Science Museum Series D (Anthropology)*, 21: 1-10.

Metspalu, M., T. Kivisild, E. Metspalu, J. Parik, G. Hudjashov, K. Kaldma, P. Serk, M. Karmin, D.M. Behar, M.T.P. Gilbert, P. Endicott, S. Mastana, S.S. Papiha, K. Skorecki, A. Torroni and R. Villems, 2005, "Most of the Extant mtDNA Boundaries in South and Southwest Asia were Likely Shaped During the Initial Settlement of Eurasia by Anatomically Modern Humans", *BMC Genetics*, 6: 26-36.

Metspalu, Mait, Gyaneshwer Chaubey, Bayazit Yunusbayev, Irene Gallego Romero, Monika Karmin, Chandana Basu Mallick, Ene Metspalu, Sadagopal Shanmugalakshmi, Karuppiah Balakrishnan, Kumarasamy Thangaraj, Lalji Singh, Ramasamy Pitchappan, Toomas Kivisild and Richard Villems, 2010, "The Genetic Structure of South Asian Populations as Revealed by 650 000 SNPs", *Genome Biology*, 11: (suppl. 1), 08.

Metspalu, Mait, Irene Gallego Romero, Bayazit Yunusbayev, Gyaneshwer Chaubey, Chandana Basu Mallick, Georgi Hudjashov, Mari Nelis, Reedik Maegi, Ene Metspalu, Maito Remm,

Ramasamy Pitchappan, Lalji Singh, Kumarasamy Thangaraj, Richard Villems, and Toomas Kivisild, 2011, "Shared and Unique Components of Human Population Structure and Genome-Wide Signals of Positive Selection in South Asia", *American Journal of Human Genetics*, 89(6): 731-44.

Oota, H., K. Kurosaki, S. Pookajorn, T. Ishida and S. Ueda, 2001, "Genetic Study of Palaeolithic and Neolithic Southeast Asians", *Human Biology*, 73: 225-31.

Oppenheimer, Steven, 2004, *The Original Eve: Modern Man's Journey Out of Africa*, New York: Carol and Graf Publishers.

Palanichamy, M.G., C. Sun, S. Agrawal, H.J. Bandelt, Q.P. Kong, F. Khan, C.Y. Wang, T.K. Chaudhuri, V. Palla and Y.P. Zhang, 2004, "Phylogeny of Mitochondrial DNA Macrohaplogroup N in India, Based on Complete Sequencing: Implications for the Peopling of Indian Subcontinent", *American Journal of Human Genetics*, 75(6): 966-78.

Petraglia, Michael, Christopher Clarkson, Nicole Boivin, Michael Haslam, Ravi Korisettar, Gyaneshwer Chauhan, Peter Ditchfield, Dorian Fuller, Hannah James, Sacha Jones, Toomas Kivisild, Jinu Koshy, Marta Mirazon Lahr, Mait Metspalu, Richard Roberts and Lee Arnold, 2009, "Population Increase and Environmental Deterioration Correspond with Microlithic Innovations in South Asia ca. 35,000 Years Ago", *Proceedings of the National Academy of Sciences of the United States of America*, 106(30): 12261-266.

Primorac, D., D. Marjanovic, P. Rudan, R. Villems and P.A. Underhill, 2011, "Croatian Genetic Heritage: Y-chromosome Story", *Croat. Med. J.*, 52(3): 225-34.

Rajkumar, R. and V.K. Kashyap, 2003, "Haplotype Diversity in Mitochondrial DNA Hypervariable Regions I and II in Dravidian Linguistic Group in India", *Forensic Science International*, 48: 27-37.

Rajkumar, R., 2004, "Haplotype Diversity in Mitochondrial DNA Hypervariable Regions I and II in Three Communities in Southern India", *Forensic Science International*, 136: 79-82.

Rajkumar et al., 2005, "Phylogeny and Antiquity of M Macrohaplogroup Inferred from Complete mtDNA Sequence of Indian Specific Lineages", *BMC Evolutionary Biology*, 5(6): 5-26.

Redd, A.J., T. Karafet, M. Bamshad, B.L. Jorde, J.M. Naidu, B. Walsh and M.F. Hammer, 2002, "Gene Flow from the Indian Subcontinent to Australia: Evidence from the Y-Chromosome", *Current Biology*, 12: 673-77.

Richards, M. and V. Macaulay, 2001, "The Mitochondrial Gene Tree Comes of Age", *American Journal of Human Genetics*, 68(6): 1315-20.

Mountain, J.L., J.M. Hebert, S. Bhattacharyya, P.A. Underhill, C. Ottolenghi, M. Gadgil and L.L. Cavalli-Sforza, 1995, "Demographic History of India and mtDNA-sequence Diversity", *Am. J. Hum. Genet.*, 56(4): 979-92.

Sahoo, S., A. Singh, G. Hemabindu, J. Banerjee, T. Sitalaximi, S. Gaikwad, R. Trivedi, P. Endicott, T. Kivisild, M. Metspalu, R. Villems and V.K. Kashyap, 2006, "A Prehistory of Indian Y-chromosomes: Evaluating Demic Diffusion Scenarios", *Proceedings of the National Academy of Sciences of the United States of America*, 103(4): 843-48.

Semino, O., G. Passarino, P.J. Oefner, A.A. Lin, S. Arbuzova, L.E. Beckman, G. De Benedictis,

P. Francalacci, A. Kouvatsi, S. Limborska, M. Marcikiae, A. Mika, B. Mika, D. Primorac, A.S. Santachiara-Benerecetti, L.L. Cavalli-Sforza and P.A. Underhill, 2000, "The Genetic Legacy of Paleolithic *Homo sapiens sapiens* in Extant Europeans: A Y-chromosome Perspective," *Science*, 290(5494): 1155-59.

Sengupta, S., L.A. Zhivotovsky, R. King, S.Q. Mehdi, C.A. Edmonds, C.E. Chow, A.A. Lin, M. Mitra, S.K. Sil, A. Ramesh, M.V. Usha Rani, C.M. Thakur, L.L. Cavalli-Sforza, P.P. Majumder and P.A. Underhill, 2006, "Polarity and Temporality of High-resolution Y-chromosome Distributions in India Identify both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists", *Am. J. Hum. Genet.*, 78(2): 202-21.

Sharma, Gunjan, Rakesh Tamang, Ruchira Chaudhary, Vipin Kumar Singh, Anish M. Shah, Sharath Anugula, Deepa Selvi Rani, Alla G. Reddy, Muthukrishnan Eaaswarkhanth, Gyaneshwer Chaubey, Lalji Singh and Kumarasamy Thangaraj, 2012, "Genetic Affinities of the Central Indian Tribal Populations", *Plos One* 7(2), e32546.

Singh, K.S., 1998, *India's Communities National Series: People of India*, New Delhi: Oxford University Press.

Soares, Pedro, Jean Alain Trejaut, Jun-Hun Loo, Catherine Hill, Maru Mormina, Chien-Liang Lee, Yao-Ming Chen, Georgi Hudjashov, Peter Forster, Vincent Macaulay, David Bulbeck, Stephen Oppenheimer, Marie Lin and Martin B. Richards, 2008, "Climate Change and Postglacial Human Dispersals in Southeast Asia", *Molecular Biology and Evolution*, 25(6): 1209-18.

Thangaraj, K., L. Singh, A.G. Reddy, V.R. Rao, S.C. Sehgal, P.A. Underhill, M. Pierson, I.G. Frame and E. Hagelberg, 2003, "Genetic Affinities of the Andaman Islanders: A Vanishing Human Population", *Curr. Biol.*, 13(2): 86-93.

Thangaraj, K., G. Chaubey, V.K. Singh, A. Vanniyarajan, I. Thasneem, A.G. Reddy and L. Singh 2006, "In Situ Origin of Deep Rooting Lineages of Mitochondrial Macrohaplogroup 'M' in India", *BMC Genomics*, 7: 136-51.

Thangaraj, Kumarasamy, Gyaneshwer Chaubey, Toomas Kivisild, Deepa Selvi Rani, Vijay Kumar Singh, Thanseem Ismail, Denise Carvalho-Silva, Mait Metspalu, L.V.K.S. Bhaskar, Alla G. Reddy, Sarat Chandra, Veena Pande, B. Prathap Naidu, Niharika Adarsh, Abhilasha Verma, Inaganti Amara Jyothi, Chandana Basu Mallick, Nidhi Shrivastava, Ragala Devasena, Babita Kumari, Amit Kumar Singh, Shailendra Kumar Dhar Dwivedi, Shefali Singh, Geeta Rao, Pranav Gupta, Vartika Sonvane, Kavita Kumari, Afsar Basha, K.R. Bhargavi, Albert Lalremruata, Arvind Kumar Gupta, Gurukamal Kaur, K.K. Reddy, A. Papa Rao, Richard Villems, Chris Tyler-Smith and Lalji Singh, 2008, "Maternal Footprints of Southeast Asians in North India", *Human Heredity*, 66(1): 1-9.

Torroni, A., H.J. Bandelt, V. Macaulay, M. Richards, F. Cruciani, C. Rengo, V. Martinez-Cabrera, R. Villems, T. Kivisild, E. Metspalu, J.R. Parik, H.V. Tolk, K. Tambets, P. Forster, B. Karger, P. Francalacci, P. Rudan, B. Janicijevic, O. Rickards, M.L. Savontaus, K. Huoponen, V. Laifinen, S. Koivumaki, B. Sykes, E. Hickey, A. Novelletto, P. Moral, D. Sellitto, A. Coppa, N. Al-Zaheri, A.S. Santachiara-Benerecetti, O. Semino and R. Sczzari, 2001, "A Signal, from Human mtDNA, of Postglacial Recolonization in Europe", *American Journal of Human Genetics*, 69(4): 844-52.

Trivedi, R., T. Sitalaximi, J. Banerjee, A. Singh P.K. Sircar, and V.K. Kashyap, 2006, "Molecular Insights into the Origins of the Shompen: A Declining Population of the Nicobar Archipelago", *Journal of Human Genetics*, 51(3): 217-26.

Underhill, P.A., 2003, "Inferring Human History: Clues from Y-chromosome Haplotypes", *Cold Spring Harb. Symp. Quant. Biol.*, 68: 487-93.

Underhill, P.A. and T. Kivisild, 2007, "Use of Y-chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations", *Annu. Rev. Genet.*, 41: 539-64.

Underhill, P.A., N.M. Myres, S. Roots, M. Metspalu, L.A. Zhivotovsky, R.J. King, A.A. Lin, C.E. Chow, O. Semino, V. Battaglia, I. Kutuev, M. Järve, G. Chaubey, Q. Ayub, A. Mohyuddin, S.Q. Mehdi, S. Sengupta, E.I. Rogaev, E.K. Khusnutdinova, A. Pshenichnov, O. Balanovsky, E. Balanovska, N. Jeran, D.H. Augustin, M. Baldovic, R.J. Herrera, K. Thangaraj, V. Singh, L. Singh, P. Majumder, P. Rudan, D. Primorac, R. Villemans and T. Kivisild, 2010, "Separating the post-Glacial Coancestry of European and Asian Y-chromosomes Within Haplogroup, R1a", *Eur. J. Hum. Genet.*, 18(4): 479-84.

Underhill, P.A., G. Passarino, A.A. Lin, P. Shen, M. Mirazón Lahr, R.A. Foley, P.J. Oefner and L.L. Cavalli-Sforza, 2001, "The Phylogeography of Y-chromosome Binary Haplotypes and the Origins of Modern Human Populations", *Ann. Hum. Genet.*, 65(1): 43-62.

Underhill, P.A., P.D. Shen, A.A. Lin, L. Jin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonne-Tamir, J. Bertranpetti, P. Francalacci, M. Ibrahim, T. Jenkins, J.R. Kidd, S.Q. Mehdi, M.T. Siegelstad, R.S. Wells, A. Piazza, R.W. Davis, M.W. Feldman, L.L. Cavalli-Sforza and P.J. Oefner, 2000, "Y-chromosome Sequence Variation and the History of Human Populations", *Nature Genetics*, 26(3): 358-61.

Wallace, Douglas C. and Antonio Torroni, 2009, "American Indian Prehistory as Written in the Mitochondrial DNA: A Review", *Human Biology*, 81(5/6) (October): 509-21.

Watkins, W.S., B.V.R. Prasad, J.M. Naidu, B.B. Rao, B.A. Bhanu, B. Ramachandran, P.K. Das, P.B. Gai, P.C. Reddy, P.G. Reddy, M. Sethuraman, M.J. Bamshad, 2005, "Diversity and Divergence Among the Tribal Populations of India", *Annals of Human Genetics*, 69: 680-95

Wells, R.S., N. Yuldasheva, R. Ruzibakiev, P.A. Underhill, I. Evseeva, J. Blue-Smith, L. Jin, B. Su, R. Pitchappan, S. Shanmugalakshmi, K. Balakrishnan, M. Read, N.M. Pearson, T. Zerjal, M. T. Webster, I. Zholoshvili, E. Jamarjashvili, S. Gambarov, B. Nikbin, A. Dostiev, O. Aknazarov, P. Zalloua, I. Tsoy, M. Kitaev, M. Mirrakhimov, A. Chariev and W.F. Bodmer, 2001, "The Eurasian Heartland: A Continental Perspective on Y-chromosome Diversity", *Proc. Natl. Acad. Sci. USA*, 98(18): 10244-49.

Zielinski, G.A., P.A. Myewski, L.D. Meeker, S. Whitlow, M. Twickler and K. Taylor, 1996, "Potential Atmospheric Impact of the Toba Mega Eruption ~71,000 Years Ago", *Geophysical Research Letters*, 23: 837-40.